# ENHANCING AUDIO PERCEPTION IN NOISY ENVIRONMENT

# AUDIO SIGNAL PROCESSING

A Thesis
Presented to
The Academic Faculty

by

Vasundhara Rawat

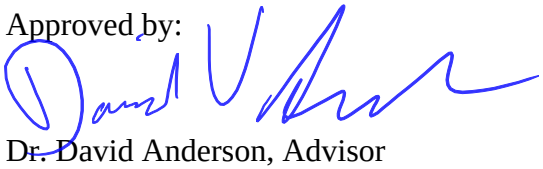
In Partial Fulfillment
Of the Requirements for the Degree
Bachelors of Science in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
May 2016

**ENHANCING AUDIO PERCEPTION IN NOISY ENVIRONMENT**

**AUDIO SIGNAL PROCESSING**

Approved by:



Dr. David Anderson, Advisor
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Dr. Mark Clements
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Date Approved: 28 April 2016

# ACKNOWLEDGEMENTS

I would like to especially thank my mother and father for their endless guidance, support, and encouragement. I would not be here without them. I would also like to thank my professor Dr. David Anderson for invaluable assistance with research design, data collection, and both personal and professional mentorship over the past semesters. I would also like to acknowledge support from both the President's Undergraduate Research Award (PURA). Finally I would also like to thank Professor Malavika Shetty for her constant guidance throughout thesis writing.

# TABLE OF CONTENTS


| | Page |
|---|---|
| ACKNOWLEDGEMENTS | iv |
| LIST OF FIGURES | vi |
| ABSTRACT | vii |
| CHAPTER | |
| 1 INTRODUTION | 1 |
| 2 BACKGROUND | 3 |
| CRITICAL BANDS AND SUBBANDS | 3 |
| HUMAN PERCEPTION OF SOUND | 3 |
| EQUIVALENT RECTANGULAR BANDWIDTH | 5 |
| 3 METHODOLOGY | 6 |
| 4 RESULTS AND DISCUSSION | 8 |
| 5 CONCLUSION | 10 |
| REFERENCES | 11 |

# LIST OF FIGURES


| | Page |
|---|---|
| Figure 1: Unwound structure of cochlea displaying the frequency mapping | 4 |
| Figure 2: Critical bands and basilar membrane | 4 |
| Figure 3: ERB showing the 23 different frequency band of the input signal | 5 |
| Figure 4: Frequency spectrum of the input (top) and output (bottom) sound | 8 |

# ABSTRACT

Enhancing audio perception in noisy environments is currently one of the most explored topics in the field of signal processing. After thorough research, engineers who have developed noise suppression filter using filter banks or any related approaches are faced with the difficulty of removing the noise caused when the input sound signal is passed through a filter. This paper introduces a technique using ERB (Equivalent Rectangular Bandwidth) that enhances the attributes of a clean signal such that it sounds clearer in the presence of background noise. The intended audiences are scientists and researchers in the field of exploring audio signal processing.

# CHAPTER 1

# INTRODUCTION

The purpose of this paper is to improve sound quality by processing the input signal so that it can be heard in the presence of environmental noise. Often when people are trying to listen to audio in a noisy environment they simply turn up the volume. While this approach might be useful in some situations, however, it has its limitations. It may not be possible to have a sharp change in volume when working with a small speaker and a battery-powered device. Moreover, turning up the volume by a significant amount would also amplify the background noise in the audio signal thereby overpowering the useful information of the signal if the signal contains background noise. Instead of providing a clearer sound it might lead to hearing loss. In order to reach an optimal solution, only the quieter parts of the speech need to be boosted. This procedure should produce a clear natural sounding speech that has been altered to be audible in the presence of background noise. An example where this would apply includes listening to an audio or phone conversation in the presence of external noise.

Acoustical background noise often accompanies mobile communication. The listener expends more effort listening, with reduced speech intelligibility due to the noisy environment, and perceives mixed far-end speech and acoustical background noise. As such, Dr. Sauert suggests enhancing the intelligibility of a clear speech signal, presented in a noise enriched environment, by implementing an algorithm to raise the average speech spectrum over the average noise spectrum, thereby regaining speech intelligibility [1].

Another approach suggested by Dr. Niederjohn in [2] and includes utilizing a high pass filter to enhance higher formants followed by rapid amplitude compression. A time-adaptive and frequency-dependent signal-to-noise ratio (SNR) recovery approach [3] is presented combined with a limitation of the spectral amplitudes to prevent hearing damage and overload of sound equipment.

Dynamic Range Compression is widely used for audio effects. This involves mapping the audio range dynamic range of an audio signal to a smaller range as described in [4]. This results in reducing signal level of higher peaks while leaving quieter parts untreated. This leads to a more regular volume for the listener and could be used to filter loud signals to prevent hearing loss without requiring the user to change the volume controls.

The approach considered in this paper is to implement an ERB (Equivalent Rectangular Bandwidth) auditory filter bank and observe the effect of implementing the filter with a gain function on each critical frequency band. The gain function implemented in this algorithm is the amplification added to each sub-band based on its amplitude in comparison to the rest of the signal. After decomposition of the input signal, higher frequencies with less energy needing additional boost for audibility will be given a larger gain. The gain in each frequency band can change quickly to accommodate changes in the signal level.

Through the above technique, the sounds of consonants will sharpen so that vocal audio may be clearly heard in the presence of external noise. The problem with modifying audio on the required time-scale is that it can result in audible distortion.

A perceptual criterion is used to allow maximal processing flexibility while eliminating or reducing perceived distortion. In addition, the total delay introduced in our algorithm was comparatively small allowing augmented listening without perceived echo.

# CHAPTER 2

# BACKGROUND

## CRITICAL BANDS AND SUBBANDS

In the paper, “A Modulation View of Audio Processing for Reducing Audible Artifacts”, Dr. Anderson suggests modulating the gain function dynamics of a signal and later describes how rapid fluctuations in gain to effectively modulate the signal results in perceptual artifacts [5]. Thus the output signal produced is not completely smooth. The paper focuses on techniques that would be used to build a system to obtain a desired signal (speech) output by applying time-varying gains without any other detectable changes. To reduce the effects of modulation, constraints are placed on the modulating gain function to either decrease the average modulation depth or to decrease the highest modulation frequency. The paper discusses ways of selecting the appropriate maximum modulating frequency and highlights the use of sub-bands.

The results of the experiments in [5] reveal that applying frequency–dependent, time–varying gain to audio may be better performed using critical-band filters rather than using constant–bandwidth sub-bands. This procedure should be followed especially if the signal varies rapidly in frequency, as it will emphasize on the important frequencies and perform operation on them individually as desired.

## HUMAN PERCEPTION OF SOUND

People hear various frequencies in a sound and the cochlea processes them differently, as shown in Figure 1, at high and low frequencies with the high frequencies

being heard through “filters” that are much wider in bandwidth than the “filters” for the low frequencies [6].

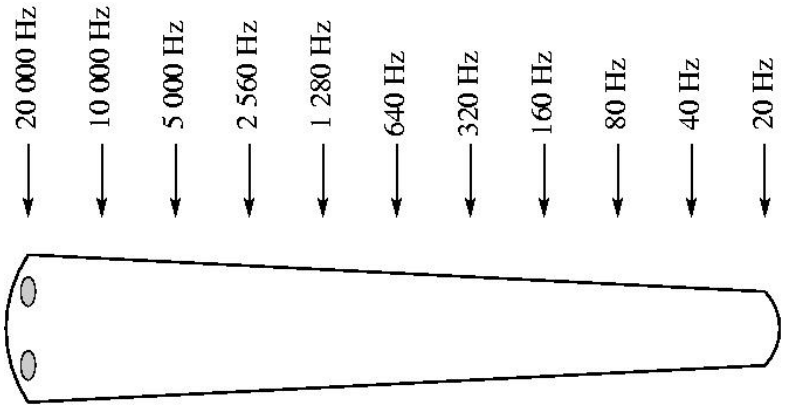



**Figure 1. Unwound structure of cochlea displaying the frequency mapping [8]**

Through Zwicker experiments [7] it is concluded that sounds with larger bandwidths sound louder. As seen in Figure 2, the critical band corresponds to a pooling along the basilar membrane: the width in terms of frequency corresponds to an estimate of the physical length along the membrane, over which auditory nerve signals are pooled. For a center frequency of 1000 Hz the critical bandwidth is 150 Hz; that corresponds to a 1.3 mm stretch along the basilar membrane. Likewise for a center frequency of 8000 Hz the critical bandwidth is 800 Hz, but this frequency range also corresponds to 1.3 mm along the basilar membrane.

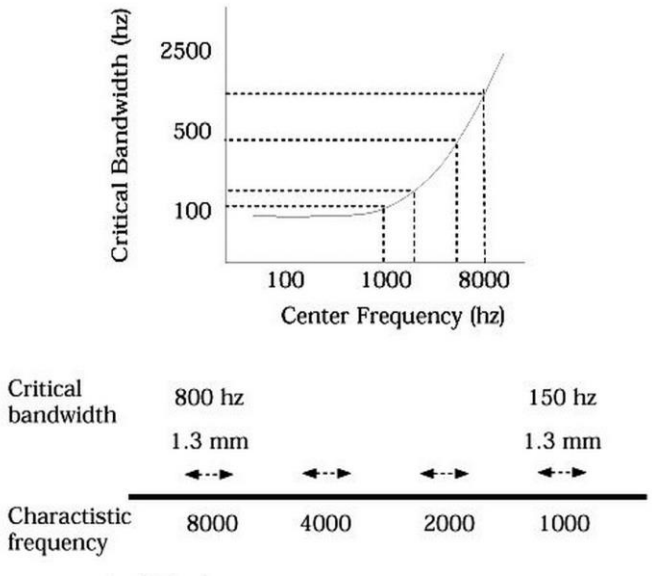



**Figure 2. Critical bands and basilar membrane [7]**

It is easier to modulate a high frequency sound than a low frequency sound by regulating the gain function. Owing to the low frequency characteristics of the signal, if the gain dynamics of the signal are changed at a rate faster than that of the frequency, then that produces roughness within the signal and defeats the purpose of achieving a perfectly smooth signal. It will lead to a really disturbing sound and might cause undesirable effects like time lag, overlapping of sound or echo problems.

## EQUIVALENT RECTANGULAR BANDWIDTH

To solve the problem of background and other additional noises, the ERB technique divides the signal into sub-bands based on frequency. The ERB also gives an approximation to the bandwidths of the filters in the cochlea, using the unrealistic but convenient simplification of modeling the filters as rectangular band-pass filters. On receiving the input signal, it should be passed to an ERB, which divides the signal into different frequency bands as shown in Figure 3, and each one of those would have a center frequency.

# CHAPTER 3

# METHODOLOGY

The filtering process includes taking each frequency band and based on the center frequency of the sub-band, select an appropriate gain instead of taking the signal as a whole and applying the gain using just one frequency threshold.

The approach considered includes the following steps:

- Take the input signal and pass it into an ERB. The ERB then splits the signal into 23 different frequency bands. This enables us to use the high and low frequency regions of the signal separately.

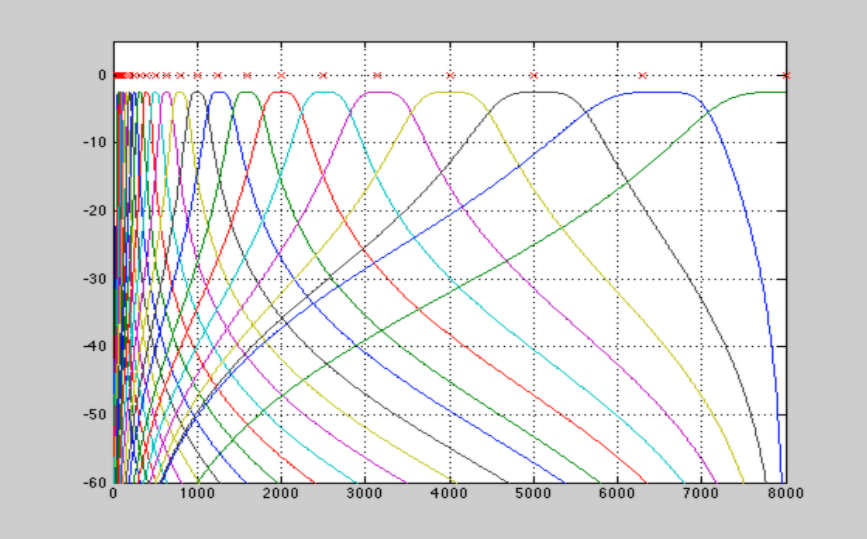



**Figure 3. ERB showing the 23 different frequency band of the input signal**

- The envelope of the signal is then calculated using Hilbert Transforms and normalized by setting the maximum gain to unity. The estimated envelope is then used as an input to the gain function designed to achieve the desired result. The envelope acts to create a mathematical approximation for further signal processing.

- The gain function is calculated using a threshold and comparing each band's envelope value to that threshold. The maximum value of the input signal is also found and the threshold is set to three-fourths of that value. If the value of the sub band is less than the product of the maximum gain and the threshold then the gain for that sub band is amplified.
- To prevent huge difference in gain values between two consecutive blocks, if the sub band signal does not qualify for amplification then its gain is set to that of that previous sub band signal.
- Once these steps have been performed, each sub-band of the signal is then applied to their corresponding gain. After the application of gain the frequency bands are combined together to produce the output signal.
- Later, noise was added to both the input signal (unprocessed) and the output signal (processed) to test which one sounded better.

# CHAPTER 4

# RESULTS AND DISCUSSION

As mentioned before, the signal is first broken down into 23 frequency bands. And then the envelope of the signal is extracted using Hilbert Transform. The envelope is then used to calculate the amplification factor corresponding to the frequency band. Figure 4 illustrates this process for one of the sub-bands.

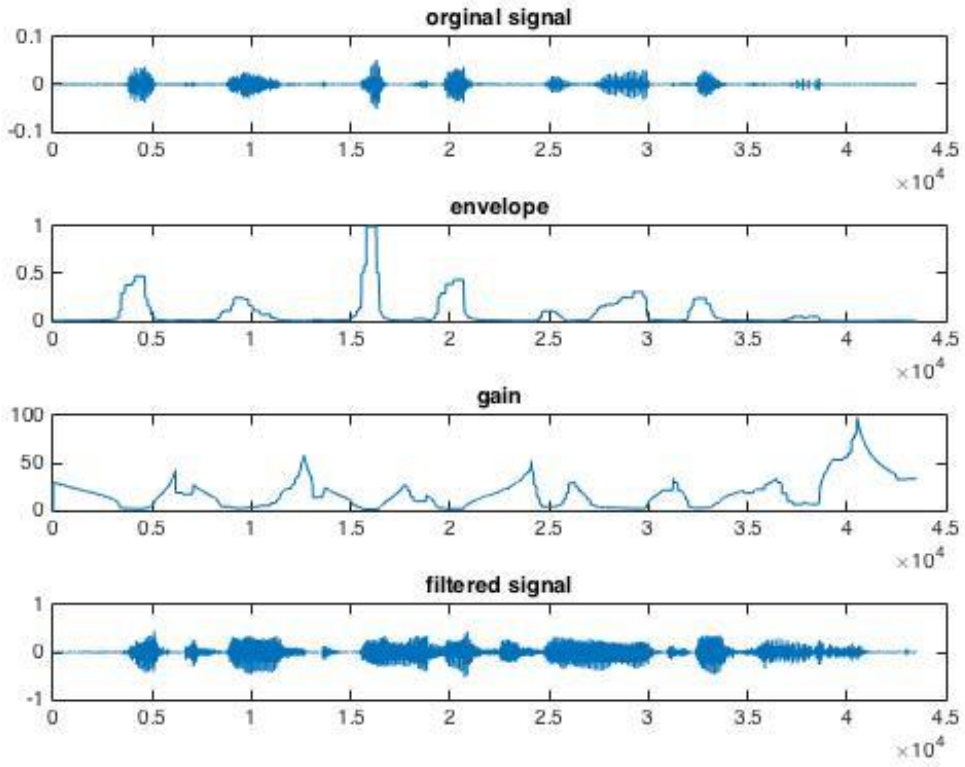



**Figure 4. The process of obtaining the filtered signal from the original signal**

The results obtained can be seen in Figure 4. To generate this plot a test input signal was in a relatively noise free environment and then was processed via the

algorithm. After the signal was processed, the spectrogram for both the signals was generated for comparison.

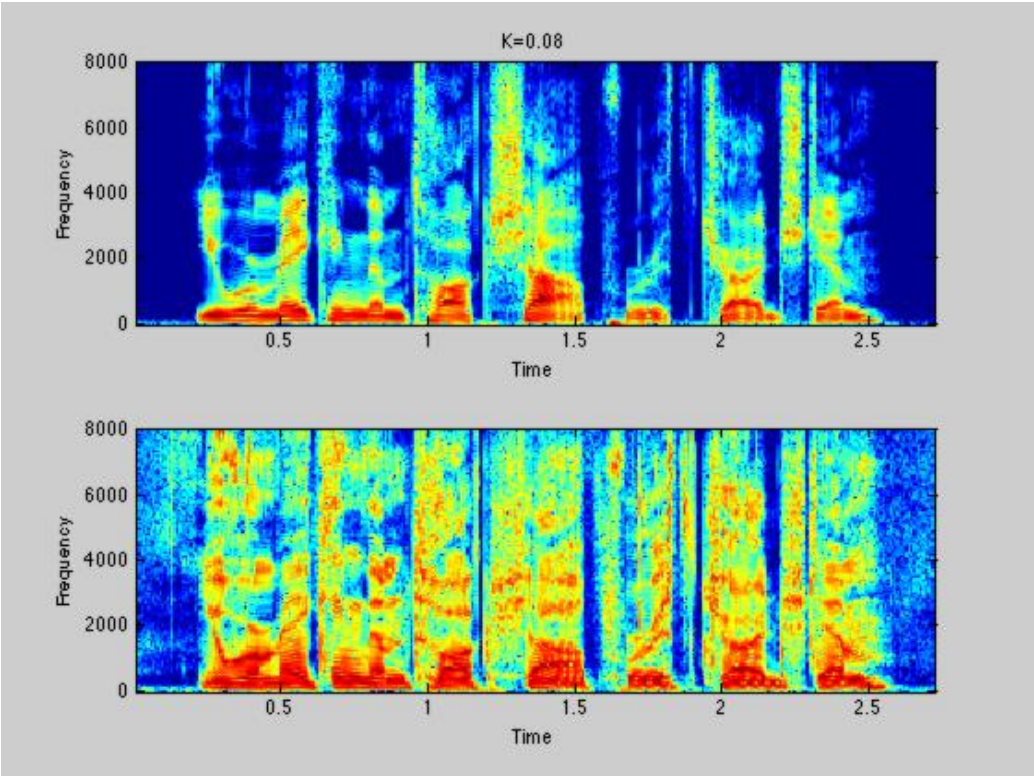



**Figure 5. Frequency spectrum of the input (top) and output (bottom) sound**

The top plot represents the frequency response of the output of an unprocessed signal and the bottom plot represents the frequency response of the output signal after being processed. The y-axis represents the frequency and the x-axis represent the sampling time. The blue region represents low-energy portions of the spectrum and the red represents high-energy portions of the speech spectrum. The yellow, green and aqua regions, low energy regions are relevant portions of the speech spectrum that need to be boosted before listening in noise.

In comparison to the unprocessed signal, we can see that the output plot of the processed signal has more prominent yellow and red traces that imply that the signal has

more boost in terms of audibility. Hence the algorithm is able to process the input signal consisting and outputs a crisper and a more audible output signal by performing frequency depended gain to the input signal.

# CHAPTER 5

# CONCLUSION

Currently the output sound corresponding to the processed signal is more audible compared to the unprocessed sound in the presence of noise yet there are some speech samples in which despite obtaining better audibility on processing, it was hard understand the whole signal in the presence of speech. This indicates that there were parts the signal that the algorithm missed and hence they didn't get enhanced properly after processing. Hence the algorithm needs further improvement.

Presently a fixed value was set for the threshold and all the speech samples were tested with the same threshold. Coming up with a generalized threshold function that changes value depending on the quality of the signal passed in would produce better results when testing with diverse audio samples.

Another aspect where the algorithm might not be as effective would be in the case if the audio signal passed in to process contains too much noise already. In this case the algorithm might amplify some of the noise contents of the signal thereby making it sound coarser instead of crisper. Hence some kind of noise suppression algorithm implementation to take care of the noise present in the input signal would be beneficial.

# REFERENCES


[1] B. Sauert and P. Vary, “Near end listening enhancement: Speech intelligibility improvement in noisy environments,” in IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), vol. I, May 2006, pp. 493–496.

[2] Russell J. Niederjohn and James H. Grotelueschen, “The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression,” in Proc. of ICASSP, Aug. 1976, vol. 24, pp. 277–282.

[3] Russell J. Niederjohn and James H. Grotelueschen, “Speech intelligibility enhancement in a power generating noise environment,” in Proc. of ICASSP, Aug. 1978, vol. 26, pp. 378–380.

[4] D. Giannoulis, M. Massberg and J. D. Reiss, “A Tutorial on Digital Dynamic Range Compressor Design,” J. Audio Eng. Soc., vol. 60, pp. 399–408 (2012 June).

[5] David V. Anderson, “A modulation view of audio processing for reducing audible artifacts,” in 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Mar. 2010, number 1, pp. 5474–5477, IEEE.

[6] Brian C. J. Moore, An Introduction to the Psychology of Hearing, Academic Press Limited, 4th edition, 2003.

[7] D. Heeger, “Perception Lecture Notes: Loudness Perception and Critical Bands.”*Perception Lecture Notes: Loudness Perception and Critical Bands*. Department of Psychology, New York University.

[8] “Physics 224 Lecture 6.” *Physics 224 Lecture 6*, Web.